European Journal of Human Genetics (2000) 8, 154–156

www.nature.com/ejhg

MEETING REPORT

Discovery, scoring and utilization of human single nucleotide polymorphisms: a multidisciplinary problem

Anders Isaksson[1], Ulf Landegren[1,2], Ann-Christine Syvänen[2], Peer Bork[3], Christian Stein[4], Flavio Ortigao[5] and Anthony J Brookes[6]

[1]Department of Genetics and Pathology, Rudbeck Laboratory; [2]Department of Medical Sciences, Molecular Medicine, Uppsala University Hospital, Uppsala, Sweden; [3]European Molecular Biology Laboratory, Heidelberg; [4]Fraunhofer Patent Centre, Patent and Licensing Agency for DHGP, Munich; [5]Interactiva Biotechnologie GmbH, Ulm, Germany; [6]Center for Genomics Research, Karolinska Institute, Stockholm, Sweden

There are great hopes that the most common form of human genetic variation, single nucleotide polymorphisms (SNPs), can be used to improve radically biological understanding and to advance medicine. However, considerable controversy exists over just how SNPs can be applied to gain these insights. The second international SNP meeting, held at Schloss Hohenkammer, Munich, Germany, brought together leading international scientists from academia and industry to look at these issues from a multidisciplinary perspective. Topics that were covered spanned SNP discovery, scoring technologies, population genetics, disease studies, commercial dimensions, pharmacogenomics, bioinformatics, and legal considerations. SNP discovery is picking up speed; The SNP Consortium (TSC) is set to produce 300 000 publicly available SNPs within 2 years. Improved technologies for scoring SNPs are reducing hands-on time and cost, although truly high-throughput methods are still lacking for genome-wide population-based studies. Large numbers of SNPs have already been analysed in diverse populations. The results emphasise the importance of considering population history when using SNPs to search for genetic risk factors. Opinions on the feasibility of extensive SNP-based analysis of complex disease vary. However, combining expertise from several fields will be key to achieving optimal utilization of SNPs. *European Journal of Human Genetics* (2000) **8**, 154–156.

Keywords: single nucleotide polymorphisms; association studies; complex disease; genotyping methods

Introduction

SNPs have evoked tremendous excitement because of the notion that they might allow identification of genes associated with complex diseases. However, how the growing body of information about human genetic variation can best be used for this purpose remains unclear. The idea of using populations of cases and controls for association studies is especially appealing, since such samples are far easier to obtain than the family materials used in conventional linkage analyses. The ultimate goal of being able to perform whole genome association studies using hundreds of thousands of SNPs in ultra-high throughput assays is currently far from being realised, due to the present deficit of SNPs, and the lack of appropriate methods. However, both these obstacles may be about to be overcome. Other factors must also be considered in the search for SNPs associated with complex disease, such as the history of the investigated population and the confounding issues of allelic and genetic heterogeneity.

Correspondence: Anthony Brookes, Center for Genomics Research, Karolinska Institute, Theorells väg 3, S-171 77 Stockholm, Sweden.
Tel: +46 8 728 6630; Fax: +46 8 331 547;
E-mail: Anthony.Brookes@cgr.ki.se
Received 20 October 1999; revised 1 November 1999; accepted 11 November 1999

SNP discovery

The accumulation of SNPs in the private sector, limiting research access, provoked ten large pharmaceutical companies, five academic genome centres and the UK Wellcome

Trust to join forces to create The SNP Consortium (TSC). As summarised by Arthur Holden (TSC Chairman, Chicago, USA), this enterprise has launched a program to discover and publish 300 000 random genetic SNPs within two years. Christian Stein (Frauenhofer Patent Centre, Munich, Germany) pointed out that there is a distinct trend that patents on partial nucleic acid sequences or SNPs will only be granted in Europe, when it is possible to describe a function and commercial application. This is in agreement with the HUGO Statement on Intellectual Property, soon to be released, which was presented by Gert-Jan van Ommen (President of HUGO, Netherlands).

Which among the millions of SNPs in the human genome are most likely to prove useful for association studies? SNPs located close to functionally important genes are more valuable as markers than random genomic SNPs. Moreover, SNPs located in coding (cSNPs) or regulatory (rSNPs) regions have the added benefit of potentially being the genetic variation directly contributing to disease. Ken Buetow (National Cancer Institute, Maryland, USA), Shamil Sunyaev (EMBL, Heidelberg, Germany), and Don Morris (Incyte Pharmaceuticals, Palo Alto, USA) all reported on efficient discovery of SNPs by automated alignment and comparison of EST sequences. The exercise is highly cost effective, and the intragenic location of these variants could make them more useful on an SNP by SNP basis than the markers that will be found by TSC. Several databases are becoming available that will offer ready access to the growing number of SNPs. Marianne Siegfried (Interactiva GmbH, Germany, and Karolinska Institute, Sweden) presented the HGBASE SNP database, which has a human gene focus. Heikki Lehväslaiho (EMBL-EBI, Cambridge, UK) described software and database solutions to link locus-specific mutation databases with SNP databases and other resources.

SNP scoring technologies

Two principal approaches are taken to meet the demands for high throughput, low cost, and highly accurate SNP scoring. The first approach is to develop fast and simple methods for scoring SNPs in individual reactions in microtiter format. Dynamic Allele-Specific Hybridization (DASH) assay (Anthony Brookes, Karolinska Institute, Stockholm, Sweden) monitors oligonucleotide probe hybridisation using a low-cost fluorescent intercalating dye over a temperature gradient to give accurate allele discrimination. DNA-polymerase-assisted allele distinction is utilised in pyrosequencing (Pål Nyrén, Royal Institute of Technology, Stockholm, Sweden) and in the template-directed dye-terminator (TDI) method (Pui-Yan Kwok, Washington University School of Medicine, St Louis, USA). The signal detection in pyrosequencing is based on luminescence induced by pyrophosphate released during the primer extension reaction, while the TDI-assay uses fluorescence polarisation, an inexpensive detection principle. An alternative to PCR-based methods, the Invader assay, was presented by Shaun Lonergan (Third Wave Technologies, Inc., Wisconsin, USA). This assay is based on accumulation of cleavage products from a junction structure between pairs of probes and a genomic target sequence.

The other principal approach to high-throughput, low-cost genotyping is to score multiple SNPs from each sample in a multiplexed fashion. Ann-Christine Syvänen (Uppsala University, Sweden) and Tomi Pastinen (National Public Health Institute, Helsinki, Finland) used minisequencing and allelic specific extension on microarrays to determine the frequencies of disease-causing mutations in the Finnish population. Minisequencing (single-base extension) reactions in solution were combined with a genetic array of oligonucleotides carrying sequences complementary to tag sequences on the primers for multiplex SNP scoring on DNA microarrays (Pamela Sklar, Whitehead Institute Genome Center, Cambridge, USA), and on fluorescence-labelled microspheres (Allen Roses, Glaxo Wellcome, North Carolina, USA). Because of problems connected with multiplex PCR, the requirement of a PCR-step prior to the detection reaction is the major bottleneck in methods for multiplex SNP scoring. Analysis of SNPs with the aid of replication of padlock probes on microassays (Anders Isaksson, Uppsala University, Sweden), may offer a future means to avoid the requirement for target amplification by multiplex PCR.

Population genetics

Beside the availability of abundant SNP markers and efficient means to score them, a number of other factors will influence the success of attempts to locate disease genes via SNPs. Critical questions include over what distance pairs of allelic genetic variants can be expected to be in linkage disequilibrium (LD), that is alleles at nearby loci tend to occur in specific combinations. The history of an investigated population might be important, as a population having recently undergone a severe bottleneck may exhibit LD over extended distances, since less recombination has taken place. The extent of LD will influence the number of SNPs that need to be analysed. In addition, founder populations may exhibit less allelic heterogeneity–a lower number of variants of any gene that may contribute to a particular phenotype.

Alison Dunning (University of Cambridge, Cambridge, UK) reported work to establish LD maps in four human populations that differed in the time elapsed since they were founded. She observed LD between most neighbouring markers and sometimes between markers over 100 kb apart. Ryk Ward (University of Oxford, Oxford, UK) also stressed the need to establish and evaluate population histories when using SNP markers, and he pointed out the value of converting SNP data to haplotype data that represent what are the observed combinations of alleles along segments of chromosomes. In this regard, Svante Pääbo (Max Plank Institute for Evolutionary Anthropology, Leipzig, Germany) described his group's work to characterise the haplotypes

found in an approximately 10 kb region of low recombination frequency on the human X chromosome. The study was also extended to encompass other primates, revealing the far greater sequence diversity among chimpanzees, and to identify haplotypes ancestral to mankind.

Also in the population genetic session, Charles Aquadro (Cornell University, New York, USA), proposed using SNP analyses to look for footprints of adaptive evolution. He suggested that analyses of ETLs – evolutionary trait loci – could serve to identify regions that include alleles which have become targets of selective fixation. These regions could include beneficial alleles that have come to dominate in the population, and could therefore be important targets of study.

Diseases of phenotypes

There are three general strategies for using SNPs in association studies of complex diseases:

i) SNPs are used to home-in further on disease-related mutations within cM regions previously identified by genome linkage scans;

ii) genetic variations are screened in and around candidate genes to implicate them in disease and ultimately to identify significant variants, and

iii) most ambitiously, genome wide association studies may be attempted using extensive sets of SNPs.

A small number of successful association studies continue to spur researchers on, since it is thereby known that these strategies can sometimes work. Allen Roses (Glaxo Wellcome, North Carolina) showed that a 4 MB regional SNP map around the APOE locus would have enabled one to locate the increased risk that the E4 allele of this gene contributes to Alzheimer's disease. But this risk allele is in many ways rather an ideal case (10-fold increased risk in homozygotes and medium frequency allele, with similar effect in almost all populations studied). His group now attempts to narrow down chromosomal regions previously identified by linkage in psoriasis, migraine, and type II diabetes.

A 'pathway-based' candidate gene approach to Alzheimer's disease was illustrated by Anthony Brookes. By spreading an investigation over four biochemical processes comprising 250 candidate genes and 400 intragenic SNPs, the relative etiologic importance of each pathway is starting to become clear. Ryk Ward presented detailed single-gene analyses. He is evaluating the correlation between *ACE* gene haplotypes and indices of blood pressure. The data provide a nice example of how estimations of haplotype configurations and cladistic analyses can lead towards the precise intragenic location of the pathogenic allele(s).

Joseph Terwilliger (Columbia University, New York, USA), pointed out that in order to establish association between marker SNPs and a disease phenotype two crucial links must hold. The marker must be significantly associated with the disease allele, and the disease allele in turn must have a clear-cut relation to the disease phenotype. Only if both these associations are sufficiently strong can the marker be connected to the disease. The first of these links can be strengthened by using more markers, considering haplotypes and taking population history into account. The second association depends on the relative importance of the disease allele. It is therefore of outstanding importance to use a suitable population and to enrich for disease alleles over environmental risk factors in the ascertainment process. This matter became the primary focus of an almost 2 hour discussion session.

Conclusion

A combination of academic and commercial efforts is driving SNP research forward at an increasing pace. As a result, a considerable fraction of the human genetic variation will be known in the near future and concomitant efforts are underway to translate this knowledge into medical advances. However, this work will have important consequences also in other areas. For instance, it will be possible to obtain a more detailed view on human migration and history, and on the genetic variation that distinguishes humans from closely related species. The use of SNPs to investigate genetic factors influencing normal variation in properties such as behaviour, is likely to cause controversy. We will therefore have to find acceptable ways to manage this form of genetic research at the community level.